International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-2, Issue-2, March-2014

Review on Detecting DDoS Attacks using Map Reduce in Haddop

Mr.Akshay Dattatray Shete, Mr.Sudhanshu S. Gonge

***Abstract*— An assault on a network that overflows it with so many requests that regular traffic is either decelerated or entirely interrupted. Unlike a virus or worm, this can cause severe damage to databases. A Distributed Denial of service (DDos) attack can employ hundreds or even thousands of computers that have been previously flooded by HTTP GET packet. The massive amounts of data that collect over time which problematic to analyze using common database management tools. Big data includes activity logs (machine generated data) which consist of unstructured format capture from web. The repository is continuously challenged as Big data increases exponentially where security is one of the challenging and harmful concerns. To handle Big data, Hadoop technology takes cardinal part in analysis. I have proposed detection of DDos attack by using Counter based algorithm and Access Pattern algorithm which will implement using Map Reduce in Hadoop framework. Besides, we are predicting future behavior of attacker by means of analytical & statistical results.**

***Index Terms*— DDoS, HDFS, UDF, Algorithm.**

I. INTRODUCTION

A Denial of service (DoS) attack is way to make a computer resource unavailable to normal users. The Dos attacks are becoming more powerful due to bot behavior recently. Attack that forces numerous sources to create the denial-of-service state is known as The Distributed Denial of Service (DDos) attack. DDos attacks are enormous hazard to internet services. HTTP flooding attack is one of the distinctive DDos attack in that hosts are directing bulky quantity of request to target website to expend its resources[1,3]. Currently, there is immense growth in internet traffic. Due to this many DDos attack detection systems facing a problem. DDos attack detection systems are categorized into two parts software based systems & hardware based system. According to paper proposed by Jinghe Jin & team they implemented the hardware based system for DDos attack detection. One of the major advantages of hardware-based DDos defense systems is that they can route packets at a greater speed. But problem with this systems are high false positive rate. The way to avoid this problem of false positive rate and big data traffic I am compelling to use software based system grounded on Hadoop platform through which we can solve problem of high traffic rate. In this paper it is going to discuss Hadoop Framework can use in terms to address the problem of big data which is caused by DDos attacks [2]. It is going to be discussed counter based algorithm to detect DDos attack & solve the problem of high false positive rate and Access pattern based algorithm to detect DDos attack using behavior of attacker [1, 2].

In this paper above mentioned DDos detection algorithms along with predictive analysis using Revolution analytics ('R') and GUI will be designed using dashboard tool since it supports organized business with meaningful & useful data.

Figure 1 shows the overview of proposed system. In that workflow of system is explained. There are six basic steps data collection, data storage, data parsing & preprocessing, data processing, data analytics, data visualization. Data collection & storage is done by flume tool. In Hadoop architecture data is preprocessed & parsed. Data processing is done by two algorithms counter based algorithm & access pattern based algorithm. In data analytics I am predicting future behavior of attacker. Finally all the results are shown on dashboard [1,2].

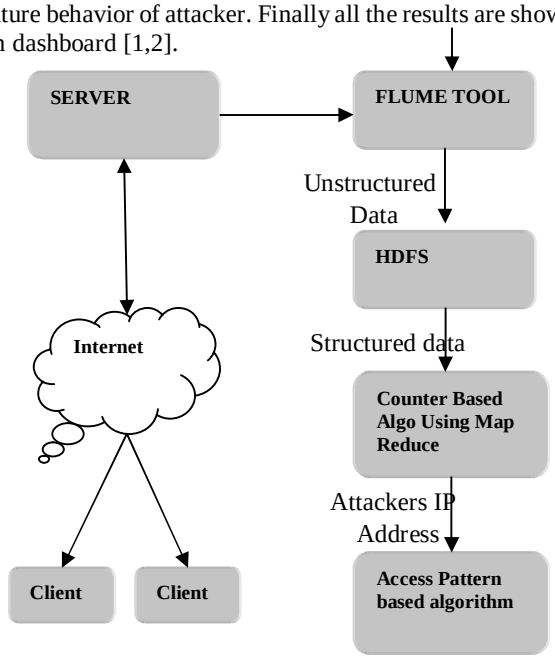


Fig 1: Counter Based Algorithm

Mr.Akshay Dattatray Shete, Third year student of I.T. Department, KJ's,Trinity College of Engineering and Research,Pune-48, Pune, India, 9561978898., (E-mail: akshay.s0890@gamil.com).
Mr.Sudhanshu S. Gonge, Assist. Professor of I.T. Department, KJ's, Trinity College of Engineering and Research, Pune-48, Pune, India, 9766578019. (E-mail: gongesudhanshu@gmail.com).

2. DDOS ATTACK DETECTION

2.1 Data Extraction & Storage

Flume's architecture is simple, robust, and flexible in our system, we are using it for data extraction & storage. It collects log data from a set of application servers.

The deployment consists of a number of logical nodes, arranged into three tiers architecture. The first tier is the agent tier. Agent nodes are typically installed on the machines that generate the logs and which is data's initial point of contact with Flume. They forward data to the next tier of collector nodes, the collectors then aggregate the streams into larger streams which can be written efficiently to a storage tier such as HDFS [3,4].

2.2 Counter Based Algorithm In Pig Latin (Data Processing)

Counter-based detection is a simple method that counts the total track volume or the number of web page requests by manipulating machine generated log file. Since the DDos attack with the low volume of trace such as the HTTP GET incomplete attack is predominant in these days, the frequency of page requests from clients will be a more dynamic factor. Map Reduce algorithm to detect DDos with URL counting. To lower the false positive rate, we adopted response rate against page requests as secondary regulation as well as trace volume, which was proposed by Liu et al.

The Figure2 depicts algorithm prerequisites three input parameters of time interval, threshold and unbalance ratio, which can be loaded through the distributed cache mechanism of Map Reduce. Time interval limits monitoring duration of the page request[4].Threshold specifies the permitted frequency of the page request to the server against the prior normal status, which regulates whether the server should be warned or not. The unbalance ratio variable represents the anomaly ratio of response per page request between a particular client and a server. This value is used for picking out attackers from the clients. Algrithm generates key values of server IP address, masked timestamp, and client IP address[4,5,7].

The time interval is used for counting the number of requests from a specific client to the specific URL within the same time duration. Finally it summarizes the number of URL requests, page requests, and server responses between a client and a server. When total requests for a specific server exceeds the threshold, the Map Reduce job emits records whose response ratio against requests is greater than unbalance ratio, marking them as attackers. It needs a prerequisite to know the threshold value from historical monitoring data in advance.We can analyse previous logs for compute the average of request count, this computed value will be treated a threshold[1,6,8].

We can use Java programming for implementation of counter based algorithm but it involves worst space complexity.So we can reduce this by using Pig script.It uses SQL statements to deal with data and in the background it runs java bytecode.This definitly helps to improve space complexity and execution in distributed environment[7,8].

3. DATA ANALYTICS

3.1 Access Pattern Algorithm Using 'R'

The access pattern-based detection method assumes that clients reveal the symptom of bot behavior and that attackers could be segregated from normal clients. Through R analytics obtains access sequence to the web page between a client and a web server and calculates the spending time according to time interval of each request of the URL. Besides, it searches infected hosts by comparing the access sequence and the spending time among clients trying to access the same server web page address. R provides statistical terminology which will need to predicate the attacker.

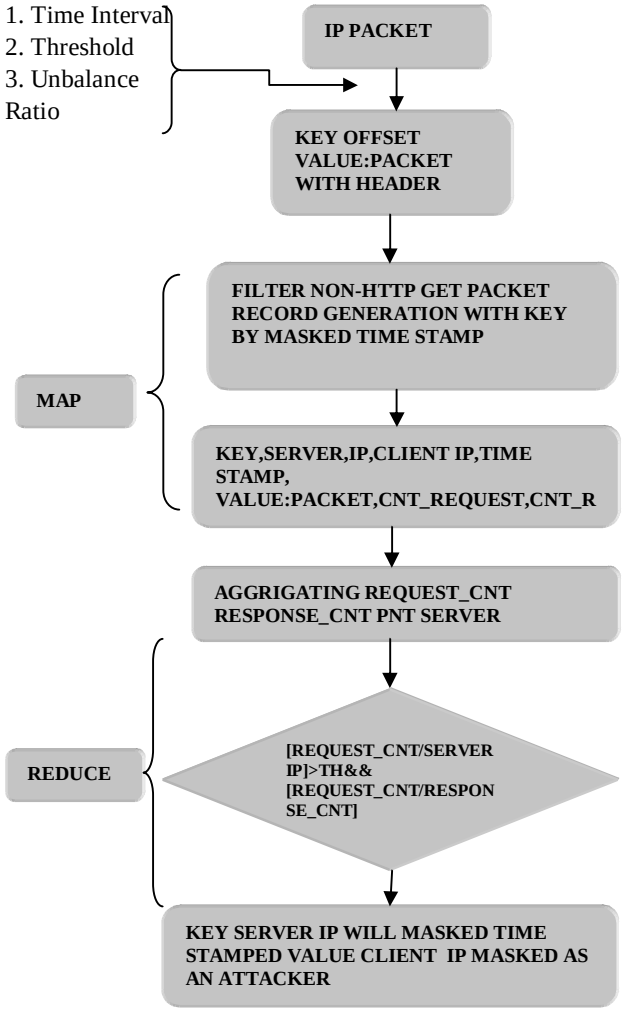


Fig 2: Counter Based Algorithm

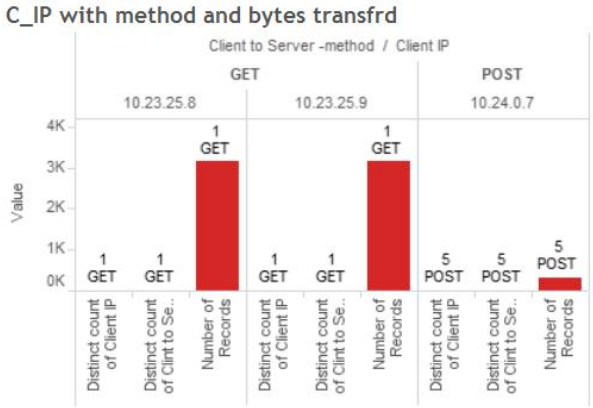
Revolution Analytics addresses opportunities in Big Data Analytics while supporting the following objectives for working with Big Data Analytics:

1. Optimizing business process and reducing operational cost.
2. Reducing the risk by anticipating and mitigating problems before they occur
3. Reduce data moment and replication along with optimize computational speed. Traditional IT infrastructure is simply not able to meet the demands of new "Big Analytics" landscape. For these reasons, many enterprises are turning to the

"R" statistical programming language and Hadoop (both open source projects) as a potential solution to this solve the commercial need. That's why we are implementing Second Access pattern based algorithm in 'R'.

4. DATA VISUALIZATION

Dashboards give signs about a business letting you know something is wrong or something is right. The corporate world has tried for years to come up with a solution that would tell them if their business needed maintenance or if the temperature of their business was running above normal. [1,8]*Dashboards* give you an overview of how your properties are performing by displaying summaries of different reports as widgets on a single or multiple pages through hyper link. It is difficult to collect, review, or analyze massive amounts of data can be overwhelming for anyone.That's why industry leaders such as Microsoft and IBM devote so many resources to developing dashboard technology and dashboard software.



In this paper it is proposed, a modern dashboard over more conventional data collection and visualizing methods such as manual recording or non-real-time manual input software. We are going to enactment the output of server which is in the format of log file using the modern dashboards technique unexpected results are reported. Because replication is required for scientific progress, papers submitted for publication must provide sufficient information to allow readers to perform similar experiments or calculations and use the reported results. Although not everything need be disclosed, a paper must contain new, useable, and fully described information. For example, a specimen's chemical composition need not be reported if the main purpose of a paper is to introduce a new measurement technique. Authors should expect to be challenged by reviewers if the results are not supported by adequate data and critical details. Provided by Tableau Public software [8]. Tableau provide excellent feature over the traditional software such as it brings your data to life with interactive graphs, charts and maps that will engage your readers. With a few clicks you can embed your interactive graphs, dashboards, maps and tables anywhere and share with everyone. It provides tremendous graphical statistic information in fraction of second which will helpful to dish out genuine user and attacker [8].

II. CONCLUSION

In this paper, It has proposed that the system which consists of implementation Counter based and Access pattern algorithm by using Map Reduce in Hadoop. With this we can use analytics to predict the future behaviour of attacker. The better user interface provided by means of Dashboard.

REFERENCES

[1] S. Byers, A. D. Robin and D. Kormann. "Defending against an internet-based attack on the physical world". ACM Transactions on Internet Technology, 4(3): 239-254, August 2004.

[2] Jinghe Jin, NazarovNodir, Chaetae Im , SeungYeob Nam," Mitigating http get flooding attacks through modified net fpga reference router" 1-st Asia NetFPGA Developers Workshop, June 13–14, 2010, Daejeon, Korea

[3] http://en.wikipedia.org/wiki/Apache_Hadoop

[4] T. White, Hadoop: The Definitive Guide. O'Reilly Media, Inc., USA, 2009

[5] Hae-Duck J. Jeong, WooSeok Hyun*, Jiyoung Lim, and IlsunYou "Anomaly teletraffic intrusion detection systems on hadoop-based platforms survey of some problems and solutions" 2012 15th International Conference on Network-Based information system

[6] Yeonhee Lee and Youngseok Lee "Detecting ddos attacks with hadoop", ACM Student Workshop, December 6 2011, Tokyo, Japan

[7] http://www.tableue.com/dashboards